Update on the beta ensembles

Brian Rider (Temple University)

with M. Krishnapur (IISC), J. Ramírez (Universidad Costa Rica), B. Virág

(University of Toronto)

The Tracy-Widom law(s)

Consider a random Hermitian $n \times n$ matrix M with centered independent entries. The most basic fact(s) about the asymptotic spectrum is that the counting measure of (the normalized by $\frac{1}{\sqrt{n}}$) eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ satisfies

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k}(\lambda) \to \frac{1}{2\pi}\sqrt{4-\lambda^2}\, d\lambda \quad \text{(Wigner semi-circle)},$$

and $\lambda_{max}, \lambda_{min} \to \pm 2, a.s.$

In the mid-90's Tracy-Widom computed the fluctuations of λ_{max}, in the complex Gaussian case:

$$\mathbb{P}\left(n^{2/3}(\lambda_{\max} - 2) \leq t\right) \to \exp\left(-\int_t^{\infty} (s-t)u^2(s)ds\right)$$

in which u solves $u''(t) = tu(t) + 2u^3(t)$ (Painlevé II) with $u \sim Ai$ at $+\infty$.

Orthogonal polynomials and gap formulas

The essential fact in the business is that the joint density of eigenvalues (on $\mathbb{R}^n$) is proportional to:

$$\prod_{k=1}^{n} e^{-\frac{1}{2}n\lambda_k^2} \times \prod_{j<k} |\lambda_j - \lambda_k|^2 \propto \det\Big(K_n(\lambda_i, \lambda_j)\Big)_{1\le i,j\le n}$$

where K_n is the kernel of the projection operator onto the span of the (first n) Hermite polynomials.

In particular the eigenvalue process is *determinantal*: all finite dimensional correlations are determinants of the same kernel function. That is,

$$\int_{\mathbb{R}^{n-k}} \det\Big(K_n(\lambda_i, \lambda_j)\Big)_{1\le i,j\le n} d\lambda_{k+1} \cdots d\lambda_n = C_{n,k} \det\Big(K_n(\lambda_i, \lambda_j)\Big)_{1\le i,j\le k}.$$

Moreover, for any such determinantal process it holds

$$\mathbb{P}\Big(\text{no points in } B\Big) = \det_{L^2(B)}(I - K_n).$$

Airy kernel and back to Painlevé

Using this “gap formula”, the limit law takes the form:

$$\mathbb{P}\left(n^{2/3}(\lambda_{\mathsf{max}} - 2) \le t\right) \to \det_{L^2[t,\infty)}(I - K_{Airy}) := F_2(t)$$

where

$$K_{Airy}(x,y) = \frac{Ai(x)Ai'(y) - Ai(y)Ai'(x)}{x - y}.$$

This is a statement about the behavior of the OPs in the vicinity of their largest zero, and follows from passing the limit “under” the determinant (the convergence takes place in trace norm.

More good fortune stems from the fact that the Airy kernel is “integrable”, implying that the resolvent kernel takes the same Christoffel-Darboux form. The opening move to the Painlevé expression is the basic formula

$$\frac{\partial}{\partial t} \log\det(I - K) = -\mathrm{tr}\left((I - K)^{-1}\frac{\partial K}{\partial t}\right).$$

Real and quaternion ensembles

Replacing the complex Gaussians with real or quaternion Gaussians, the eigenvalue density is changed as in

$$\prod_{j<k} |\lambda_j - \lambda_k|^2 \text{ is replaced } \prod_{j<k} |\lambda_j - \lambda_k|^1 \text{ or } \prod_{j<k} |\lambda_j - \lambda_k|^4.$$

And we speak of $\beta = 1, 2,$ or 4 ensembles (real, complex, or quaternion).

When $\beta = 1$, 4, the eigenvalue processes are Pfaffian (not determinantal): everywhere one had a determinant of a scalar kernel before there is now a Pfaffian of a skew matrix kernel.

Still there exist limit laws F_1 and F_4 for λ_{max} in terms of Painlevé II:

$$F_1(t) = \exp\left(-\frac{1}{2}\int_t^\infty u(s)ds\right) F_2^{1/2}(t),$$

$$F_4(t/\sqrt{2}) = \cosh\left(\frac{1}{2}\int_t^\infty u(s)ds\right) F_2^{1/2}(t).$$

Universality and ubiquity

Two directions of universality (in terms of RMT proper):

"General" Wigner matrices: Soshnikov (2000), Peché-Sosnhikov (2007), Tao-Vu (2010), Lee-Yin (2012), Bourgade-Erdös-Yau (2013).

"General" unitary/orthogonal ensembles, replacing $e^{-\text{tr} M^2} dM$ with $e^{-\text{tr} V(M)} dM$ resulting in general families of orthogonal polynomials in the correlation kernels: Deift-Gioev (2007), Pastur-Shcherbina (2003), Shcherbina (2009).

Tracy-Widom everywhere:

Longest increasing subsequence,

Last passage percolation,

Flux in TASEP/ASEP,

The whole KPZ craze.

Beta ensembles

For any $\beta > 0$, consider the law P_β on n points with density a multiple of

$$\prod_{k=1}^{n} e^{-\frac{\beta}{4} n \lambda_k^2} \times \prod_{j<k} |\lambda_j - \lambda_k|^\beta,$$

giving you back the matrix ensembles already discussed for $\beta = 1, 2, 4$.

2007, J. Ramírez, B. Virág and I introduced the "general beta Tracy-Widom laws" as the distributional limit of the largest point under P_β. It can be defined via: with b a Brownian motion,

$$TW_\beta = \sup_{f \in \mathcal{L}} \frac{2}{\sqrt{\beta}} \int_0^\infty f^2(x) db(x) - \int_0^\infty [(f'(x))^2 + x f^2(x)] dx.$$

Here $\mathcal{L}$ are those functions with $\int_0^\infty f^2 = 1, \int_0^\infty [(f')^2 + x f^2] < \infty$ and $f(0) = 0$.

This proved a conjecture of Edelman-Sutton.

Tridiagonals

Set for any $\beta > 0$,

$$H_\beta = \frac{1}{\sqrt{n\beta}} \begin{bmatrix} g_1 & \chi_{(n-1)\beta} & & & \\ \chi_{(n-1)\beta} & g_2 & \chi_{(n-2)\beta} & & \\ & \ddots & \ddots & \ddots & \\ & & \chi_{2\beta} & g_{n-1} & \chi_\beta \\ & & & \chi_\beta & g_n \end{bmatrix}$$

where each g (Gaussian) χ_r (a Chi) is independent. Then, the eigenvalues of H_β have P_β as their joint density function.

Due to Dumitriu and Edelman: at $\beta = 1, 2$ can be arrived at by the well known Householder transformations.

The variational characterization says: $-TW_\beta$ is the ground state eigenvalue for

$$\mathcal{SAO}_\beta = -\frac{d^2}{dx^2} + x + \frac{2}{\sqrt{\beta}} b'(x),$$

with Dirichlet conditions at the origin.

Really we prove that there is a coupling so that $n^{2/3}(2I - H_\beta) \to \mathcal{SAO}_\beta$ in strong resolvent norm.

$\mathcal{SAO}_\beta$ - *technical aside*

For this thing to make sense need to control the white noise by the "good" (linear) part of the potential (plus H^1 norm). We prove the inequality: for all $f \in \mathcal{L}$,

$$\left| \int_0^\infty f^2 db_x \right| \le c \int_0^\infty [(f')^2 + x f^2] \, dx + C(c,\omega) \int_0^\infty f^2 dx,$$

with $C < \infty$ almost surely.

Rests on slightly clever integration-by-parts: writing $b_x = \overline{b}_x + (b_x - \overline{b}_x)$ for $\overline{b}_x = \int_x^{x+1} b_y dy$, have for instance

$$\int_0^\infty f^2 d(b_x - \overline{b}_x) \le \int_0^\infty (f'(x))^2 dx + \int_0^\infty f^2(x) \left(\int_x^{x+1} (b_y - b_x) dy \right)^2 dx$$

and while $b_x \sim x^{1/2+}$ as $x \uparrow \infty$, the increment $\sup_{|y-x|\le 1}(b_y - b_x)$ only grows like $\log x$.

(So need nothing as fast as linear to control the white noise.)

Some immediate questions

1. Why are some values of beta special?

1'. Or, really, are any values of beta special?

2. Where is Painlevé? (Can you get formulas?)

3. Is $\mathcal{SAO}_\beta$ a rich enough characterization of the laws? (Meaning, say, can it be used as a tool to prove universality?)

PDE/diffusion descriptions

For a fixed λ consider a solution ψ of

$$SAO_\beta \psi(t) = \lambda \psi(t), \quad \psi'(0) = w\psi(0), \text{ and set } X(t) = X(t,\lambda) = \frac{\psi'(t)}{\psi(t)}.$$

This is a diffusion:

$$dX(t) = \frac{2}{\sqrt{\beta}} db(t) + (\lambda + t - X^2(t))dt, \quad X(0) = w,$$

and

$$F(\lambda, w) := \mathbb{P}_w\Big(t \mapsto X(t) \text{ never explodes}\Big) = \mathbb{P}_{w,\lambda}\Big(t \mapsto X(t,0) \text{ never explodes}\Big),$$

is the unique (up to a normalization) nonnegative bounded solution od

$$\frac{\partial F}{\partial \lambda} + \left(\frac{2}{\beta}\right)\frac{\partial^2 F}{\partial w^2} + (\lambda - w^2)\frac{\partial F}{\partial w} = 0.$$

Fact: $F(\lambda, \infty) = \mathbb{P}(TW_\beta < \lambda)$. We have no direct proof of the PDE formulation.

Amplification/application: spiked matrices

Consider the Gaussian sample covariance (or Wishart) matrices of the form XX^T for X an $n \times m$ matrix of independent Gaussians. The scaling limit of λ_{max} is still Tracy-Widom.

An honest (statistical) question is whether the largest eigenvalue can sense changing the population covariance from null. That is, take instead the ensemble $X\mathbf{\Sigma}X^T$, and, to make life easier, take $\mathbf{\Sigma}$ to be the identity save for the 11-entry some $c > 0$.

In 2005 Baik-Ben Arous-Peché (in the case $\beta = 2$) found a phase transition:

If $c < \mathfrak{c}$: $\mathbb{P}\Big(\sigma_n(\lambda_{\text{max}} - \mu_n) \le t\Big) \to F_2(t)$.

If $c > \mathfrak{c}$: $\mathbb{P}\Big(\sigma'_n(\lambda_{\text{max}} - \mu'_n) \le t\Big) \to \int_{-\infty}^{t} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}$.

If $c = \mathfrak{c} - wn^{-1/3}$: $\mathbb{P}\Big(\sigma_n(\lambda_{\text{max}} - \mu_n) \le t\Big) \to F(t,w) = F_2(t) f(t,w)$ where f can again be described in terms of Painlevé II.

Stochastic Airy with Robin boundary conditions

The (null) Wishart matrices also have a β-tridiagonalization. Now the model is BB^T where B is given by:

$$B = \begin{bmatrix} \chi_{m\beta} & \chi_{(n-1)\beta} & & & \\ & \chi_{(m-1)\beta} & \chi_{(n-2)\beta} & & \\ & & \ddots & \ddots & \\ & & & \chi_{(n-m+2)\beta} & \chi_{\beta} \\ & & & & \chi_{(n-m+1)\beta} \end{bmatrix}.$$

If you spike once, the bidiagonalization goes through with the only change that the 11-entry above is multiplied by $\sqrt{c}$.

Bloemendal-Virág (2011) showed that, in the critical window, the operator limit of these matrices is SAO_β with the condition $f'(0) = wf(0)$ - rather than $f(0) = 0$ - so exactly connected to the PDEs from two slides up.

Can check the BBP formulas solve this PDE. The $\beta = 4$ case was worked out (on the analytic side) by D. Wang, and completed with the help of $\mathcal{SAO}_\beta$. There is now a $\beta = 1$ formula (due to M.Y. Mo) that has not even been checked!

Universality for tridiagonals

Say you have

$$\mathcal{H} = -\frac{d^2}{dx^2} + y'(x), \qquad \text{with integrated potential } y(x) = \int_0^x \eta_y dy + \int_0^x \sigma_y db_y.$$

(SAO_β specifies $y(x) = \frac{1}{2}x^2 + \frac{2}{\sqrt{\beta}} b_x$.)

Say you also have some family of random Jacobi matrices which you put in the form:

$$H_n = -\triangle_n + \text{tridiag}\Big(y_{n,1},\, y_{n,2} \Big),$$

in which $\triangle_n$ is the second-difference operator on the space-scale δ_n (that you pick). Then...

"Theorem" *If, along with certain tightness conditions, it holds that*

$$\sum_{k=1}^{[x/\delta_n]} y_{n,1}(k) + 2y_{n,2}(k) \Rightarrow y(x),$$

then the bottom k *eigenvalues/eigenvectors of* H_n *tend to those of* $\mathcal{H}$ *in distribution.*

Those tightness conditions

Need a "drift plus noise decomposition":

$$\sum_{\ell=1}^{k} y_{n,i}(\ell) = \left(\delta_n \sum_{\ell=1}^{k} \eta_{n,i}(\ell)\right) + w_{n,i}(k), \quad i = 1, 2,$$

along with a tight sequence κ_n for which

$$\frac{1}{\kappa_n}\eta(x) - \kappa_n \le \eta_{n,1}(x/\delta_n) + \eta_{n,2}(x/\delta_n) \le \kappa_n \eta(x) + \kappa_n,$$

and

$$|w_{n,1}(x/\delta_n) - w_{n,1}(y/\delta_n)|^2 + |w_{n,2}(x/\delta_n) - w_{n,2}(y/\delta_n)|^2 \le \kappa_n(1 + \eta(x)^{1-}),$$

for all x, y with $|x - y| \le 1$.

For β-Hermite, $\eta(x) = x$, $\delta_n = n^{-1/3}$, $\eta_{n,1}(k) = 0$, and $\eta_{n,2}(k) = 1 - \frac{E\chi_{\beta(n-k)}}{\sqrt{n\beta}}$, etc.

(*) Functional convergence of the potential only needed up to indices of order δ_n^{-1}, while tightness conditions imposed on the full matrix.

(*) Only need to control the square-increment of the noise in terms of the drift.

Tridiagonals for general log-gases

Moving to a general potential beta ensemble: the joint density,

$$c\,\mathrm{exp}\left\{-n\beta\sum_{j=1}^{n}V(\lambda_j)\right\}\prod_{j<k}|\lambda_j-\lambda_k|^{\beta},$$

can be realized by the eigenvalues of $T_n = T_n(A, B)$, the symmetric tridiagonal matrix with $T_{i,i} = A_i$ for $i \le n$ and $T_{i,i+1} = T_{i+1,i} = B_i$ for $i \le n-1$ sampled from the density:

$$c'\,\mathrm{exp}\left\{-n\beta\left[\mathrm{tr}V(T_n(a,b))-\sum_{k=1}^{n-1}(1-k/n-1/(n\beta))\log(b_k)\right]\right\}.$$

The verification is the same as in Dumitriu-Edelman (for β=Hermite, etc.).

$\mathcal{SAO}_\beta$ *universality*

(Krishnapur-R-Virág) *Let* V *be a uniformly convex polynomial and define the random Jacobi matrices* T_n *as above.*

There are constants $\gamma, \mathcal{E}$ (depending only on V) *so that*

$$\gamma n^{2/3}(\mathcal{E} I_n - T_n) \to \mathcal{SAO}_\beta,$$

in the following sense: for every k*, the bottom* k*th eigenvalue and corresponding eigenvector (as an element of* L^2*) converge.*

(There is an additional constant $\vartheta = \vartheta(V)$ *so that* $\mathcal{E} I_n - T_n$ *can be viewed as acting on* $\mathbb{R}^n \subset L^2(\mathbb{R}_+)$ *with coordinate vectors* $e_j = (\vartheta n)^{1/6} \mathbf{1}_{[j-1,j](\vartheta n)^{-1/3}}$*).*

The gist of the proof

We need to show a CLT for the running sum of the field $k \mapsto (A, B)$ under the law

$$ce^{-n\beta H} da db, \quad \text{where } H = H(a,b) = \mathrm{tr}(V(T)) - \sum_{k=1}^{n-1} (1 - k/n - 1/(n\beta)) \log(b_k).$$

While the A_k and B_k are no longer independent variables, assuming V is polynomial provides a Markov field property: variables with indices that are more that $\deg V/2$ apart are conditionally independent given the variables in between.

This provides some nice intuition, but the proof is not “dynamic”. Rather, with H inheriting the convexity of V we just Taylor expand around the minimizer.

What minimizer?

The stand in for the true minimizing path are the minimizers $(a(t), b(t))$ of the "local Hamiltonian"

$$\mathcal{H}(a,b) = W(a,b) - (1-t)\log b \quad \text{where} \quad W(a,b) = \frac{1}{\ell}\mathrm{tr}V(C_\ell(a,b)),$$

in which C_ℓ is the $\ell \times \ell$ circulant matrix with second row $(b, a, b, 0, 0, 0, \dots)$ and $\ell > \deg V$.

The equations for the minimizers are then

$$\frac{i}{2\pi}\int_{R_t}^{L_t} \frac{sV_t(s)}{\sqrt{(s-L_t)(R_t-s)}}ds = 1, \quad \int_{R_t}^{L_t} \frac{V_t(s)}{\sqrt{(s-L_t)(R_t-s)}}ds = 0.$$

where

$$V_t(x) = \frac{1}{1-t}V(x), \quad L_t = a(t) - 2b(t), \quad R_t = a(t) + 2b(t).$$

These are precisely the "moment conditions" for the limiting eigenvalue counting measure μ_{V_t}.

This identifies the edge (centering for λ_{max}) as $\mathcal{E} := \mathcal{E}(0) = R_0$.

Other universality results and "regularity"

Even at $\beta = 2$ can only expect Tracy-Widom if the limiting eigenvalue density μ_V satisfies certain conditions.

Bourgade-Erdös-Yau: $V \in C^4$, μ_V has "one band", and $\beta \geq 1$. (Uses Dyson Brownian motion.)

Bekerman-Figalli-Guionnet: $V \in C^{31}$, μ_V also "one band", but $\beta > 0$. (By transportation of measure.)

"Regularity" is implicit in all the assumptions.

Simple large deviation ideas show μ_V minimizes

$$I(\mu) = \int_{-\infty}^{\infty} V(x)\mu(dx) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \log|x-y|\mu(dx)\mu(dy).$$

For quadratic V this is the semi-circle law, regular means (in this context) that (the density) $\mu(x)$ vanishes like a square-root at its right-most point of support.

Regularity is generic, but can produce $O(x^{\frac{4k+1}{2}})$ vanishing with polynomial $V's$.

Exotic edge limits

Though we use convexity in fundamental technical ways...

For us, the regular $n^{2/3}$ scaling from the differentiability of a certain variable edge $\mathcal{E}(\epsilon)$ at $\epsilon = 0$. More generally it is known that

$$\text{If } \mu_V(t) \sim (\mathcal{E} - t)^{\frac{4k+1}{2}} \text{ as } t \uparrow \mathcal{E} \text{ then } \mathcal{E} - \mathcal{E}(\epsilon) \sim \epsilon^{\frac{1}{2k+1}} \text{ as } \epsilon \downarrow 0.$$

So, imagining that in the general case (A, B) has a Gaussian limit with a similar profile (mean/covariance connected to the variable edge as above) we are led to the conjecture that:

Assume V is nonregular with "degree" $k \geq 1$. Then there are constants γ, $\mathcal{E}$ so that

$$H_{n,k} = \gamma n^{2/4k+3}\Big(\mathcal{E} I_n - T_n(A, B)\Big)$$

converges in the sense of our main theorem to the operator

$$\mathcal{S}_{\beta,k} = -\frac{d^2}{dx^2} + x^{\frac{1}{2k+1}} + \tfrac{2}{\sqrt{\beta}} x^{-\frac{k}{2k+1}} b'_x.$$

There are Painlevé formulas at $\beta = 2$ due to Claeys-Its-Krasovsky.